Canopy biochemistry estimation using spectrodirectional CHRIS data

Huber, Silvia ; Kneubühler, Mathias ; Koetz, Benjamin ; Schopfer, Jürg T ; Zimmermann, Niklaus E ; Itten, Klaus I

Abstract: Sun and sensor geometry cause spectrodirectional effects in remotely sensed reflectance data which can influence the estimation of biophysical and biochemical variables. Previous studies indicated that vegetation indices can be strongly influenced by such effects and thus impact the results. This study examined the uncertainty induced by changing view angles on statistical methods used for nitrogen concentration (CN) estimation. We analyzed data of the spaceborne ESA-mission CHRIS (Compact High Resolution Imaging Spectrometer) on-board PROBA-1, which provides hyperspectral and multi-angular data with a spatial resolution of 17 m. The images were acquired in June 2005 over a test site in Switzerland and subsequently preprocessed. Linear regression models (LM's) were developed between laboratory-measured CN, reflectance and transformed reflectance (continuum-removed and normalized), respectively, using a subset selection algorithm. For each CHRIS observation angle a particular LM was built. All LM's were evaluated using 10-fold cross-validation with random splitting order of the data. By considering the adjusted R2 (adj.-R2), the root mean square error (RMSE) and percent error (% error), the LM's were finally compared. Best CN predictions were achieved with models calibrated on nadir data with R2 of 0.63 and 0.59 using the reflectance and transformed reflectance, respectively. Generally, better LM's were attained with nadir and -36° data than with +36° data (forward scatter direction), for both, reflectance and transformed reflectance. Applying nadircalibrated LM's to off-nadir data was not successful to estimate CN. The results suggest that the CHRIS/PROBA mission provides useful data for biochemistry estimation. Caution is required when applying statistical methods developed on nadir data to data with directional effects.

Originally published at:
Huber, Silvia; Kneubühler, Mathias; Koetz, Benjamin; Schopfer, Jürg T; Zimmermann, Niklaus E; Itten, Klaus I (2006). Canopy biochemistry estimation using spectrodirectional CHRIS data. In: 2nd International Symposium on Recent Advances in Quantitative Remote Sensing (RAQRS), Torrent, Valencia, Spain, 25 September 2006 - 29 September 2006, 314-319.

Canopy Biochemistry Estimation Using Spectrodirectional CHRIS Data

Silvia Huber[1], Mathias Kneubühler[1], Benjamin Koetz[1], Jürg T. Schopfer[1], Niklaus E. Zimmermann[2] and Klaus I. Itten[1]
[1]*Dept. of Geography, Remote Sensing Laboratories, University of Zurich, Switzerland;*
[2]*Swiss Federal Research Institute WSL, Land Use Dynamics, Switzerland*
shuber@geo.unizh.ch

ABSTRACT – *Sun and sensor geometry cause spectrodirectional effects in remotely sensed reflectance data which can influence the estimation of biophysical and biochemical variables. Previous studies indicated that vegetation indices can be strongly influenced by such effects and thus impact the results. This study examined the uncertainty induced by changing view angles on statistical methods used for nitrogen concentration (C_N) estimation. We analyzed data of the spaceborne ESA-mission CHRIS (Compact High Resolution Imaging Spectrometer) on-board PROBA-1, which provides hyperspectral and multi-angular data with a spatial resolution of 17 m. The images were acquired in June 2005 over a test site in Switzerland and subsequently preprocessed. Linear regression models (LM's) were developed between laboratory-measured C_N, reflectance and transformed reflectance (continuum-removed and normalized), respectively, using a subset selection algorithm. For each CHRIS observation angle a particular LM was built. All LM's were evaluated using 10-fold cross-validation with random splitting order of the data. By considering the adjusted R^2 (adj.-R^2), the root mean square error (RMSE) and percent error (% error), the LM's were finally compared. Best C_N predictions were achieved with models calibrated on nadir data with R^2 of 0.63 and 0.59 using the reflectance and transformed reflectance, respectively. Generally, better LM's were attained with nadir and -36° data than with +36° data (forward scatter direction), for both, reflectance and transformed reflectance. Applying nadir-calibrated LM's to off-nadir data was not successful to estimate C_N. The results suggest that the CHRIS/PROBA mission provides useful data for biochemistry estimation. Caution is required when applying statistical methods developed on nadir data to data with directional effects.*

1 INTRODUCTION

Sun and sensor geometry cause spectrodirectional effects in remotely sensed reflectance data which can influence the estimation of biophysical and biochemical variables. These effects can be seen as an additional information source. Numerous studies have shown that bidirectional measurements contain added information about vegetation structure, such as gap fraction and leaf orientation (Chen et al., 2003; Ustin et al., 2004), or tree cover and tree height (Heiskanen, 2006) and that separability of land cover types can be improved with multiangular information (Barnsley et al., 1997).

On the other hand, the anisotropic reflectance behaviour for instance of plant canopies implies that remote observations can vary without a change in the physical or chemical properties of the material observed. This makes it difficult to interpret remotely sensed data of the same geographic location collected from different instruments, spatial scales or times (Asner, 2004). The bidirectional variability is thus often considered as noise and its impact on the estimation of plant biochemical and structural variables remains unknown in many cases. For vegetation indices (VI's), which are widely used for plant properties retrieval, it has been demonstrated that they respond differently to viewing angles. Some were significantly sensitive to angular effects, others remained invariant (Gemmell and McDonald, 2000; Verrelst et al., 2006). This emphasizes the importance to test methods on their directional sensitivity because bidirectional effects exist in much of the remotely sensed data commonly used, especially for sensors with large fields of view or off-nadir viewing capabilities. In this study we focused on statistical methods to investigate the impact of bidirectional effects. We used multiple linear regression models to predict C_N from untransformed as well as transformed reflectance data, measured from different view angles. Finally, predicted C_N was compared to laboratory-measured data, the performance of LM's as a function of directional information evaluated and the uncertainty induced by changing view angles assessed.

2 MATERIALS AND METHODS

2.1 Study Site

The study site (7°53' E, 47°16'N) was located in the Swiss Plateau at an altitude of ~400 meters above see level. The forest canopy was composed of a mixture of needleleaf and broadleaf species, dominated by European beech (*Fagus sylvatica* L.), European ash (*Fraxinus excelsior* L.), black alder (*Alnus glutinosa*), silver fir (*Abies alba*) and Norway spruce (*Picea abies* L.).

2.2 Field Data Retrieval

During a two-week field campaign in summer 2004, we collected foliar material from the top of tree canopies to determine biochemistry in the laboratory and measured additional biophysical and positional tree properties. A tree climber excised leaf samples, which were sealed in bags and stored in cool environment for transportation. A total of 60 foliage samples were harvested from nine different tree species and two plant functional types in order to be analyzed for nitrogen concentration in the laboratory. To obtain representative samples, from each of three different upper sunlit canopy branches we collected 15 leaves from broadleaf species and 50–60 needles from the first three needle years from needleleaf species.

To geo-locate the sampled individual tree crowns later in the remotely sensed images, the stem position of each tree was measured with a Trimble GeoXT GPS receiver, which corrects for multipath biases. We improved the positional accuracy by recording 20 to 40 GPS measurements per stem and applying a post processing differential correction to the recorded data using the Pathfinder Office software (Trimble, 2005). The positional measurements resulted in a mean horizontal positional error among all trees of 2.5 m, whereas the mean radius of a broadleaf and a needleleaf tree crown were found to be 5.0 m and 3.0 m, respectively.

Leaf material collection and CHRIS data acquisition occurred in two different years but during the same phonological period. We assumed a stable inter-annual C_N level (Martin and Aber, 1997).

2.3 Laboratory Analyses

For all collected samples we determined the biochemical composition in the laboratory. For nitrogen analyses the samples were dried at 65 °C until a constant weight was achieved. The ground samples were injected into an elemental analyzer (NA 2500; CE Instruments, Milan, Italy). All three sub-samples per tree (three branches) were pooled for analyses and each pooled sample was analyzed twice and checked for within-sample variation. None of the samples exceeded the threshold of 3% variation of the mean between the two measurements. The measured C_N ranged from 1.00 to 2.97 with a mean of 1.68 percent by dry weight.

2.4 CHRIS Data Acquisition and Processing

This study was based on the data of the spaceborne ESA-mission CHRIS (Compact High Resolution Imaging Spectrometer) on-board PROBA-1 (Barnsley et al., 2004), which provides hyperspectral and multi-angular data. Table 1 presents data specifications of the CHRIS sensor. CHRIS data were acquired in June 2005 over a mixed forest test site in Switzerland. We obtained a set of three images, since due to pointing problems of the sensor and cloud coverage, the -/+55° view angles were discarded. Figure 1 shows that the nominal fly-by zenith angles (FZA) of the CHRIS data acquisitions do rarely represent the actual viewing geometry for the date under investigation. The actual view angle for the nadir image was for instance -3.9° in the backward looking direction.

Table 1: CHRIS data specifications for Land Mode 5.

Sampling	Image area	View angles	Spectral bands	Spectral range
17 m @ 556 km altitude	6.5 x 13 km (372 x 748 pixels)	5 nominal angles @ -55°, -36°, 0°, +36°, +55°	37 bands with 6-33nm width	447-1035 nm

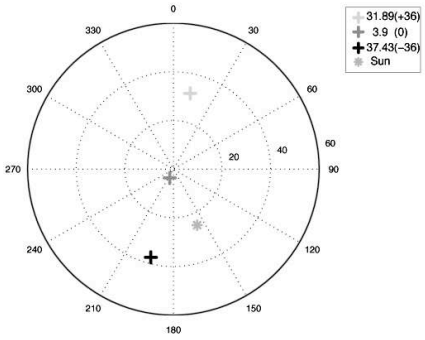

Figure 1: Acquisition geometries and illumination angles for the three CHRIS images on June 20, 2005. The nominal fly-by zenith angles are listed in brackets.

The three CHRIS images were geometrically and radiometrically corrected. Geocorrection was based on a 3D physical model (Toutin, 2004), which is implemented in the commercially available image processing software PCI/Geomatica. High positional accuracy of the respective multi-angular products after geometric correction was a prerequisite for reliable retrieval of HDRF information from the data set. The

RMSE for the specific region of interest did generally not exceed one pixel (Kneubühler et al., 2006). Subsequent atmospheric correction of the CHRIS radiance data was performed using ATCOR-3 (Richter, 1998), which is based on MODTRAN-4. ATCOR-3 accounts for terrain effects by incorporating digital elevation models (DEM) data and their derivatives such as slope and aspect, sky view factor and cast shadow. ATCOR-3 enables the processing of data from tilted sensors by accounting for varying path lengths through the atmosphere and varying transmittance. Atmospheric correction results in the generation of HDRF (Hemispherical Directional Reflectance Factor) data sets for the various CHRIS view angles.

After preprocessing of the images, the spectral signatures of the 60 field-sampled tree crowns were extracted. For further analysis two data sets were generated: the first consisted of the original spectra (termed spec) and the second of continuum-removed and normalized spectral data (termed BNC). Each data set consisted further of three images acquired from different view angles (-/+36°, 0°). In total six sets were investigated. As illustrated in Figure 2 different sensor observation angles cause significantly different reflectance. The minimum reflectance corresponds with the forward scatter direction because the sensor views the unilluminated, shadowed leaf surfaces (Sandmeier et al., 1998).

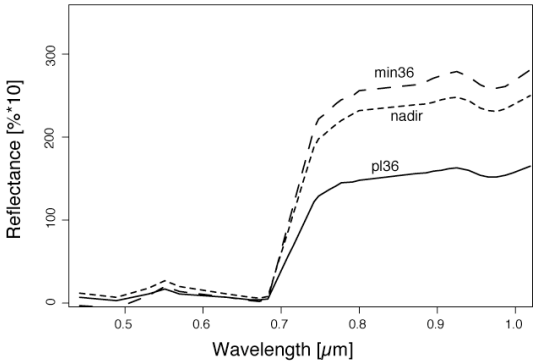

Figure 2: Spectral signatures of Norway spruce from preprocessed CHRIS data of the nominal -/+36° and nadir observation angles.

Continuum removal was developed to enhance the spectral features of interest (Clark and Roush, 1984). The observed spectral continuum is considered an estimate of the other absorptions present in the spectrum, not including the one of interest. The continuum-removed spectra were calculated by dividing the original reflectance values by the corresponding values of the continuum line (Kokaly and Clark, 1999). The continuum removal was conducted over the whole spectral range of the data by using the ENVI software (Research Systems, 2004). From the continuum-removed reflectance, the band depth of each point in the absorption feature was computed by subtracting the continuum-removed reflectance from one. To minimize extraneous influences, we applied a normalization procedure on band depths (Figure 3). The band depth normalized to the center (BNC) was calculated by dividing the band depth of each band by the band depth at the band center (Eq. 1):

$$BNC = \frac{1-(R/R_i)}{1-(R_c/R_{ic})}, \quad (1)$$

where R is the reflectance of the sample at the waveband of interest, R_i is the reflectance of the continuum line at the waveband of interest, R_c is the reflectance of the sample at the absorption feature center and R_{ic} is the reflectance of the continuum line at the absorption feature center. The band center is the

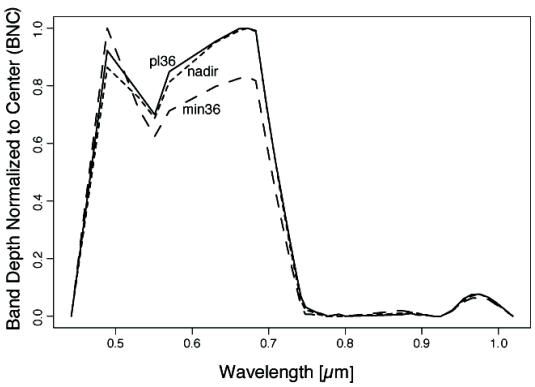

Figure 3: BNC signatures of Norway spruce from preprocessed CHRIS data of the nominal -/+36° and nadir observation angles.

minimum of the continuum-removed absorption feature (Curran et al., 2001; Kokaly and Clark, 1999). For each sample the feature center was individually calculated.

2.5 Statistical Analyses

We used multiple linear regression analyses in order to predict C_N from original spectral as well as transformed BNC data. For each data set (spec and BNC) and view angle a particular LM was developed. To reduce the number of spectral wavebands used in

the LM's, this study employed a statistical variable selection method, an enumerative branch-and-bound (b&b) algorithm (Miller, 2002). This statistical procedure selected wavebands that best explained the nitrogen concentration. The basic characteristics of b&b methods have been addressed by several papers (Furnival and Wilson 1974; Miller 2002; Mitten 1970; Narendra and Fukunaga 1977). Branch-and-bound algorithms are efficient because they avoid exhaustive enumeration by rejecting suboptimal subsets without direct evaluation and guarantee that the selected subset yields the globally best value of any criterion that satisfies monotonicity (Narendra and Fukunaga 1977). Two different stopping criteria for model selection and optimization were applied: (1) the adjusted coefficient of determination (adj.-R^2) and (2) the Bayesian information criterion (BIC). BIC tends to penalize complex models more heavily, giving preference to simpler models in selection (Hastie et al. 2001). We used the adj.-R^2 to account for differing data set sizes and number of model terms. The adj.-R^2 lowers R^2 as the number of independent variables increases and is thus always smaller than the coefficient of determination. Therefore, it is an ideal measure to compare models that include different number of observations and predictor variable combinations (Guisan and Zimmermann, 2000).

All models were evaluated using 10-fold cross-validation with random splitting order of the data (Hastie et al., 2001). Due to random splitting order we iterated each cross-validation run ten times per model in order to obtain a more robust cross-validation estimated from which we calculated the adj.-R^2, the cross-validated RMSE (CV-RMSE) and the percent error (% error). We implemented all analyses within the R statistical package, a free software environment for statistical computing and graphics (R Development Core Team, 2005) under the GNU public license.

3 RESULTS

3.1 Calibration and Cross-validated Models

In total we developed six regression models, three based on the spec data set and the other three on the BNC data set. The models proposed by the subset selection algorithm were evaluated and finally we selected a seven-term model for the spec and a five-term model for the BNC data set. In both data sets (spec and BNC), best results were attained with nadir and -36° CHRIS data (Table 2).

Table 2: Results of linear models by regressing C_N on spectral data (spec) and transformed data (BNC).

	Calibration	Cross-validation	
	Adj.R^2 (RMSE)	Adj.R^2 (CV-RMSE)	% error
LM(C_N~spec) 7-term model			
nadir	0.69 (0.321)	0.57 (0.347)	19
min36	0.67 (0.336)	0.56 (0.364)	17
pl36	0.54 (0.391)	0.42 (0.417)	21
LM(C_N~BNC) 5-term model			
nadir	0.65 (0.340)	0.54 (0.368)	16
min36	0.58 (0.381)	0.48 (0.404)	19
pl36	0.48 (0.410)	0.37 (0.432)	20

The spec models performed in general better but they consisted of two more predictor variables. Comparing RMSE with CV-RMSE shows that all LM's have similar predictive stability. BNC models have slightly smaller % errors but RMS errors are higher.

We achieved best C_N predictions with the nadir model developed from spectral data resulting in an R^2 of 0.62 and RMSE of 0.35 (Figure 4).

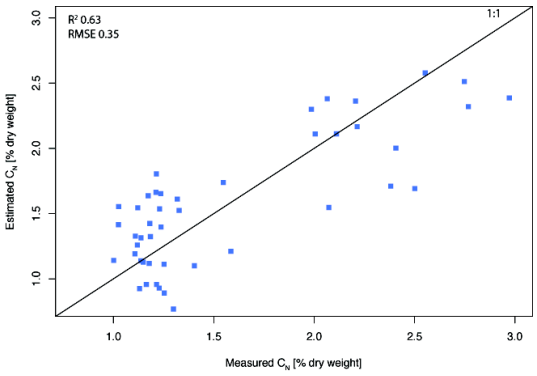

Figure 4: Relationship between measured and predicted nitrogen concentration using a linear model developed from nadir spectral reflectance (spec) data.

The five-term nadir model developed from BNC data explained 59 % of the variability in C_N with an RMSE of 0.31 (Figure 5).

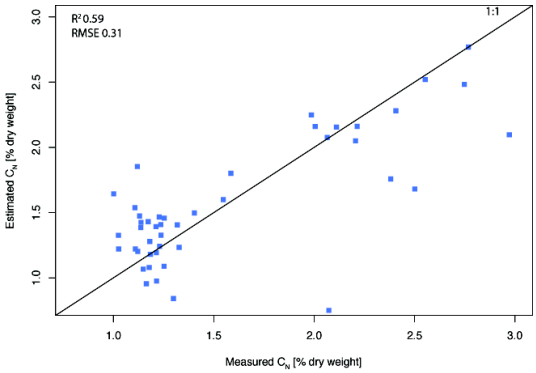

Figure 5: Relationship between measured and predicted nitrogen concentration using a linear model developed from nadir BNC data.

3.2 Uncertainty Induced by Changing View Angles

To assess the uncertainty induced by changing view angles we used models developed from nadir data to predict C_N from off-nadir data. The nadir models of both data sets (spec and BNC) predicted C_N from -36° and +36° data without success. Interestingly, models developed from nadir spectral achieved better results on +36° angles compared to -36° data, but R^2 remained still very small (0.13). Even we have chosen a less complex model for the BNC nadir data, regarding the number of terms, it predicted C_N inferior to the spec based model. The ability of a model to generalize is fundamental. As the model becomes more and more complex, it is able to adapt to more complicated underlying structures, but the estimation error increases (Hastie et al., 2001).

4 DISCUSSION AND CONCLUSIONS

In this study we applied a straightforward approach to predict C_N from multiangular CHRIS data. From field data and each observation angle of two different reflectance data sets (spec and BNC) a particular regression model was developed. The results indicate that the quality of the LM's differed among view angles but not much between data sets. LM's calibrated on nadir and -36° CHRIS data achieved higher adj.-R^2 and lower RMS Errors for both data sets. The finding that most information is contained in nadir and backwardscatter reflectance is consistent with other research which found in boreal forests an increase in bidirectional reflectance in the backscatter direction but lower forwardscatter direction reflectance (Deering et al., 1999). It was also shown that the canopy hotspot effect has rich information content for vegetation characterization, especially indications of canopy structure (i.e., a shadow is not visible) (Gerstl, 1999).

We used two different reflectance data sets (spec and BNC) to assess if bidirectional effects can be minimized with the normalization procedure. Hardly a difference was visible in the results of LM's developed from the normalized BNC and spec data set, resspectively. Angular effects were not reduced by the normalization procedure as expected. However, from the BNC data set we achieved similar results with less complex models compared to the spec data set. In spite of less complexity the generalization did not work well and thus the transfer of the BNC nadir model to off-nadir data was not successful.

A main problem of the investigation was the detection of the same pixels in all three CHRIS images, which were then assigned to field-sampled trees. This problem may be partly solved if digital surface models are used instead of a DEM for geocorrection. Uncertainties arise also from scaling point field measurements of foliar chemistry to canopy scale and comparing laboratory measurements to field concentrations because they may differ significantly (Curran, 1989). Further research might explore the combination of angular information by using the data of all five CHRIS observation angles.

ACKNOWLEDGEMENT

The authors would like to thank the Swiss National Science Foundation (SNF: project no. 200020-101517) for funding this project. The continuing effort and support of ESA and SIRA to provide CHRIS/PROBA data is gratefully acknowledged. We are very thankful to the many individuals who have helped with data collection and processing.

REFERENCES

Asner, G.P. (2004). Biophysical Remote Sensing Signatures of Arid and Semiarid Ecosystems. In S.L. Ustin (Ed.), *Remote sensing for natural resource management and environmental monitoring* (p. 53–109). Hoboken, NJ: John Wiley & Sons.

Barnsley, M.J., Allison, D., & Lewis, P. (1997). On the information content of multiple view angle (MVA) images. *International Journal of Remote Sensing, 18*, 1937-1960.

Barnsley, M.J., Settle, J.J., Cutter, M.A., Lobb, D.R., & Teston, F. (2004). The PROBA/CHRIS mission: a low-cost smallsat for hyperspectral multiangle observations of the Earth surface and

atmosphere. *Geoscience and Remote Sensing, IEEE Transactions on, 42*, 1512.
Chen, J.M., Liu, J., Leblanc, S.G., Lacaze, R., & Roujean, J.-L. (2003). Multi-angular optical remote sensing for assessing vegetation structure and carbon absorption. *Remote Sensing of Environment, 84*, 516-525.
Clark, R.N., & Roush, T.L. (1984). Reflectance Spectroscopy: Quantitative Analysis Techniques for Remote Sensing Applications. *Journal of Geophysical Research, 89*, 6329-6340.
Curran, P.J. (1989). Remote-Sensing of Foliar Chemistry. *Remote Sensing of Environment, 30*, 271-278.
Curran, P.J., Dungan, J.L., & Peterson, D.L. (2001). Estimating the foliar biochemical concentration of leaves with reflectance spectrometry testing the Kokaly and Clark methodologies. *Remote Sensing of Environment, 76*, 349-359.
Deering, D.W., Eck, T.F., & Banerjee, B. (1999). Characterization of the Reflectance Anisotropy of Three Boreal Forest Canopies in Spring-Summer. *Remote Sensing of Environment, 67*, 205.
Gemmell, F., & McDonald, A.J. (2000). View Zenith Angle Effects on the Forest Information Content of Three Spectral Indices. *Remote Sensing of Environment, 72*, 139-158.
Gerstl, S.A.W. (1999). Building a global hotspot ecology with Triana data. In, *Remote Sensing for Earth Science, Ocean, and Sea Ice Applications* (p. 184). Florence, Italy: SPIE.
Guisan, A., & Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling, 135*, 147-186.
Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: data mining, inference, and prediction*. New York: Springer.
Heiskanen, J. (2006). Tree cover and height estimation in the Fennoscandian tundra-taiga transition zone using multiangular MISR data. *Remote Sensing of Environment, 103*, 97.
Kneubühler, M., Koetz, B., Huber, S., Schopfer, J.T., Richter, R., & Itten, K. (2006). Monitoring Vegetation Growth using Multitemporal CHRIS/PROBA Data. In, *IEEE International Geoscience And Remote Sensing Symposium* (p. in Press). Denver, Colorado.
Kokaly, R.F., & Clark, R.N. (1999). Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sensing of Environment, 67*, 267-287.
Martin, M.E., & Aber, J.D. (1997). High spectral resolution remote sensing of forest canopy lignin, nitrogen, and ecosystem processes. *Ecological Applications, 7*, 431-443.
Miller, A.J. (2002). *Subset selection in regression*. Boca Raton: Chapman & Hall/CRC.
R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing. In (pp. 299-314). Vienna, Austria
Research Systems (2004). ENVI User's Guide. In: Research Systems Inc.
Richter, R. (1998). Correction of satellite images over mountainous terrain. *Applied Optics, 37*, 4004–4015.
Sandmeier, S., Muller, C., Hosgood, B., & Andreoli, G. (1998). Physical Mechanisms in Hyperspectral BRDF Data of Grass and Watercress. *Remote Sensing of Environment, 66*, 222.
Toutin, T. (2004). Review article: Geometric processing of remote sensing images: models, algorithms and methods. *International Journal of Remote Sensing, 25*, 1893-1924.
Trimble (2005). GPS Pathfinder Office Software. In: Trimble.
Ustin, S.L., Zarco-Tejada, P.J., Jacquemoud, S., & Asner, G.P. (2004). Remote Sensing of the Environment: State of the Science and New Directions. In S.L. Ustin (Ed.), *Remote sensing for natural resource management and environmental monitoring* (p. 679–729). Hoboken, NJ: John Wiley & Sons.
Verrelst, J., Koetz, B., Kneubühler, M., & Schaepman, M.E. (2006). Directional sensitivity analysis of vegetation indices from multi-angular CHRIS/PROBA data. In, *ISPRS Mid Term Symposium*. Enschede, the Netherlands.